

Bundesgesundheitsbl 2016 · 59:373–378  
 DOI 10.1007/s00103-015-2299-y  
 Online publiziert: 11. Januar 2016  
 © Die Autor(en) 2015. Dieser Artikel ist auf  
 Springerlink.com mit Open Access verfügbar



Martin Lablans<sup>1</sup> · Dennis Kadioglu<sup>3</sup> · Sebastian Mate<sup>2</sup> · Ines Leb<sup>2</sup> ·  
 Hans-Ulrich Prokosch<sup>2</sup> · Frank Ückert<sup>1</sup>

<sup>1</sup> Medizinische Informatik in der Translationalen Onkologie, Deutsches Krebsforschungszentrum,  
 Heidelberg, Deutschland

<sup>2</sup> Lehrstuhl für Medizinische Informatik, Friedrich-Alexander-Universität Erlangen-Nürnberg,  
 Erlangen, Deutschland

<sup>3</sup> Institut für medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universitätsmedizin Mainz,  
 Mainz, Deutschland

# Strategien zur Vernetzung von Biobanken

## Klassifizierung verschiedener Ansätze zur Probensuche und Ausblick auf die Zukunft in der BBMRI-ERIC

### Vernetzung von Biobanken

Moderne biomedizinische Forschung ist zunehmend auf umfangreicheres und spezielleres Biomaterial angewiesen [1]. Dieses wird in epidemiologischen Biobanken (auch populationsbezogene Biobanken genannt) oder klinischen Biobanken aufbewahrt [2]. In der deutschen Forschungslandschaft finden sich sowohl eine Vielzahl kleiner Biobanken mit nur wenigen hundert Proben – aus einer bestimmten Studie oder zu einem bestimmten (manchmal seltenem) Krankheitsbild – als auch sehr große, zentralisierte Biobanken mit hunderten tausenden von Proben, in denen z. B. die Biomaterialien aus allen Projekten eines Forschungsstandorts gemeinsam gesammelt werden. Auch wenn die neuere Entwicklung zur Etablierung entsprechender Qualitätsstandards zur Zentralisierung (und damit zu größeren Biomaterialsammlungen) beiträgt – vgl. z. B. die fünf deutschen BMBF-geförderten (Bundesministerium für Bildung und Forschung) zentralisierten Biomaterialbanken (cBMB, [3, 4]) oder die nationalen Biobanken in Estland, Island, Luxemburg und Großbritannien [5] – so stehen Forscher dennoch weiterhin vor der Herausforderung, dass die

Proben- und Datenbestände einer einzelnen Biobank nicht ausreichen, um für statistisch valide Analysen ausreichend große Kollektive zu generieren [2]. Daraus resultiert häufig die Suche nach Kooperationspartnern, die in gleichem Maße molekular gut charakterisiertes Biomaterial (inklusive der zugehörigen Probandendaten) für ähnliche Probandenphänotypen zu einem multizentrischen Projekt beisteuern können.

Die sog. „Vernetzung von Biobanken“ soll daher die Suche nach Kooperationspartnern mit geeignetem Biomaterial (oft spricht man allgemein von der „biobankenübergreifenden Probensuche“) unterstützen, ohne dass die Proben selbst dafür zentralisiert gelagert werden müssen. Sie kann sowohl auf Ebene eines einzelnen Landes – z. B. koordiniert über den Deutschen Biobankknoten („German Biobank Node“, GBN, [6] – als auch europaweit – koordiniert durch die seit 2014 etablierte paneuropäische Biobankeninfrastruktur BBMRI-ERIC („Biobanking and BioMolecular Resources Research Infrastructure-European Research Infrastructure Consortium“, [7] – etabliert werden. Auch außerhalb Europas gibt es Initiativen. Als Beispiel sei der „Biospecimen Resource Locator“

des US-amerikanischen „National Cancer Institute“ (NCI) genannt [8].

Zur IT-technischen Umsetzung einer solchen Vernetzung gibt es eine Vielzahl von Ansätzen und Strategien mit jeweils eigenen Anwendungsfällen sowie Vor- und Nachteilen. Unseren Vorschlag für eine projektunabhängige Klassifikation bestehender Strategien zur Vernetzung von Biobanken zeigt **Abb. 1**. Er unterscheidet 3 Kriterien: a) Die Granularität der die Biobanken/Proben beschreibenden Daten, b) den Speicherort derselben und c) den Grad an Automatisierung, mit dem Suchverfahren ausgeführt werden. Mittels dieser Kriterien werden bestehende Ansätze im Folgenden vorgestellt. Anschließend wird ein Konzept umrissen, das diese Ansätze in einer integrierten IT-Plattform nutzbar macht und so Forschern ein Werkzeug zur Suche nach geeignetem Biomaterial auf europäischer Ebene an die Hand geben soll.

### Strategien zur Vernetzung

Allen Vernetzungsstrategien ist gemein, dass die Probensuche keine Zentralisierung des Materials selbst erfordert. Vielmehr läuft sie auf den die Proben charakterisierenden Merkmalen

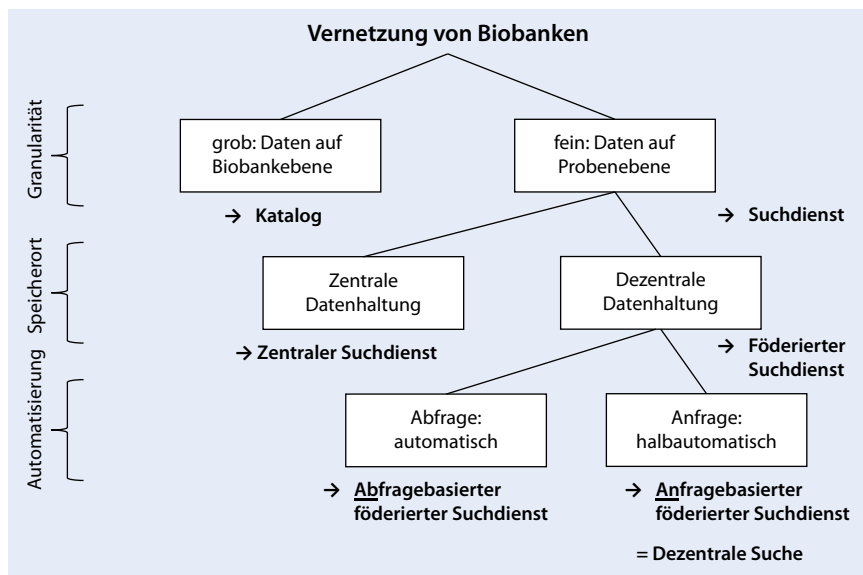


Abb. 1 ▲ Klassifikation von Ansätzen zur Suche nach Forschungspartnern mit passendem Biomaterial

sowie auf den die Probenspender charakterisierenden klinischen Daten ab. Folglich müssen die Daten, die Biobanken und ihre Proben beschreiben, durchsuchbar bzw. auffindbar gemacht werden. Ein erstes Entscheidungskriterium hierfür stellt die Granularität der Biobank-/Probanden dar:

**Kataloge.** Kataloge wie das Deutsche Biobankenregister [9] oder der BBMRI-Katalog [10] stellen einen ersten Überblick über verfügbare Biobanken auf Basis aggregierter Daten (nicht zu verwechseln mit den später noch beschriebenen Metadaten) bereit. Sie sind üblicherweise mit einem Webbrowser öffentlich einsehbar und werden daher auch mit „Schau fenstern“ [11] oder „gelben Seiten“ [12] verglichen. Datengrundlage sind meist manuelle Dateneingaben der Biobankbetreiber [13], die bereits in der Vergangenheit zur Voraussetzung für eine BMBF-Förderung gemacht wurden [3]. Neben den Adress- und Kontaktdaten einer Biobank wird als aggregiertes Datenelement z. B. die (grobe) Gesamtanzahl der für einen bestimmten Materialtyp gelagerten Proben gespeichert. Analog kann man die Probenanzahl auch mit Blick auf die relevante Krankheit der Probenspender aggregieren (z. B. Anzahl der Gewebeproben von Patienten mit einem Prostatakarzinom).

Die über einen Katalog verfügbaren Daten beinhalten also nur diejenigen „Daten auf Biobankebene“, die vom Biobankbetreiber – etwa im Rahmen einer regelmäßigen Umfrage des Katalogbetreibers oder über die Erfassungsmasken einer Webseite [13] – jeweils manuell bereitgestellt wurden. Kataloginhalte umfassen somit keinerlei Informationen zu einzelnen Proben bzw. deren Probenspendern („Daten auf Probenebene“). Diese Eigenschaft ist Vor- und Nachteil zugleich: Vorteil ist die sehr niedrige technische Hürde zur Teilnahme, denn die Registrierung im Katalog erfordert nur wenige Minuten und benötigt lediglich einen Webbrowser. Nachteil ist die potentiell mangelnde Aktualität und Granularität der manuell erfassten Daten auf Biobankebene. Eine feingranulare Suche mit Bezug zu einzelnen Proben bzw. Probenspendern, die mehrere Merkmale miteinander kombiniert, können Kataloge nicht bereitstellen. Je nach Anwendungsszenario haben derartige Kataloge sicherlich ihre Berechtigung, sollten aber aus Sicht der Autoren anderweitige Vernetzungsvarianten nur ergänzen. Für einen Vergleich bestehender Biobankkataloge sei an dieser Stelle auf [14] verwiesen.

**Zentrale Suchdienste.** Möchte man eine auf Proben-/Spendermerkmalen basierende Suche durchführen, so gibt

es wiederum zwei unterschiedliche Ansätze: zentrale Suchdienste und föderierte Suchdienste. Beispiele für zentrale Suchdienste sind CRIP [15] oder der oben genannte „Specimen Resource Locator“ des NCI [8]. Forscher können eine web-basierte Maske verwenden, um Kriterien als Schlüssel-Wert-Paare zu definieren und diese per Und/Oder-Verknüpfungen zu kombinieren. Der Suchdienst vergleicht den so entstandenen Kriteriensatz mit den ihm bis dahin an eine zentrale Datenbank übertragenen Proben-/Spenderdaten und zeigt dem Forscher sofort die Anzahl passender Proben an, ggf. auch unter Nennung derjenigen Biobanken, die entsprechende Proben gemeldet haben. Datengrundlage für diese Suche sind regelmäßige Übermittlungen strukturierter Daten, die jede Biobank aus ihrem Laborinformationssystem (LIMS) extrahieren, in das vom Suchdienst vorgegebene Zielformat transformieren und anschließend, meist über das Internet, in seine zentrale Datenbank laden muss. Dieser sog. ETL-Prozess (Extraktion, Transformation, Laden), auch materialisierte Datenintegration genannt [16], ist sowohl für die Daten übermittelnden Biobanken als auch für die entgegennehmenden zentralen Suchdienste mit der Bereitstellung entsprechend komplexer IT-Anwendungen verbunden und stellt oftmals eine nicht zu unterschätzende Hürde dar.

**Föderierte Suchdienste.** Föderierte Suchdienste arbeiten ebenfalls mit Daten auf Probenebene, verzichten aber auf das Etablieren einer zentralen Datenbank, an die die Probanden übermittelt werden müssen. Stattdessen wird eine lokale Softwarekomponente unter der Hoheit der jeweiligen Biobank betrieben, die – entweder durch materialisierte (ETL-Prozess) oder durch virtuelle Datenintegration (Adapter) – Zugriff auf die lokalen Datenbestände der Biobank erhält. Der Forscher formuliert für seine Proben-suche, ähnlich wie bei einem zentralen Suchdienst, über eine Webseite seinen Kriteriensatz durch Verknüpfung einzelner Merkmale. Die Antwort des Suchdienstes unterscheidet sich jedoch je nach Grad seiner Automatisierung, sodass zwei Klassen föderierter Suchdienste zu unter-

M. Lablans · D. Kadioglu · S. Mate · I. Leeb · H.-U. Prokosch · F. Ückert

## Strategien zur Vernetzung von Biobanken. Klassifizierung verschiedener Ansätze zur Probensuche und Ausblick auf die Zukunft in der BBMRI-ERIC

### Zusammenfassung

**Hintergrund.** Nicht selten benötigt ein medizinisches Forschungsvorhaben mehr biologisches Material, als in einer einzigen Biobank verfügbar ist. Daher unterstützt eine Vielzahl von Strategien das Auffinden potentieller Forschungspartner mit passenden Proben, auch ohne dass diese zuvor in einer zentralisierten Sammlung zusammengeführt werden müssen.

**Ziel.** Der vorliegende Beitrag beschreibt die Klassifizierung verschiedener Strategien zur Vernetzung von Biomaterialbanken, speziell zur Probensuche, sowie eine IT-Infrastruktur, die diese Ansätze kombiniert.

**Material und Methoden.** Bestehende Strategien lassen sich nach drei Kriterien klassifizieren: a) Granularität der Proben-

daten: grobe Daten auf Bankebene (Katalog) vs. feingranulare Daten auf Probenebene, b) Speicherort der Probandaten: zentrale (zentraler Suchdienst) vs. dezentrale Datenhaltung (förderierte Suchdienste) und c) Automatisierungsgrad: automatisch (abfragebasiert, förderierter Suchdienst) vs. halbautomatisch (anfragebasiert, dezentrale Suche). Alle genannten Suchdienste setzen eine Datenintegration voraus; dabei helfen Metadaten bei der Überwindung semantischer Heterogenität.

**Ergebnisse.** Der „Common Service IT“ in BBMRI-ERIC („Biobanking and Biomolecular Resources Research Infrastructure-European Research Infrastructure Consortium“) vereint einen Katalog, die dezentrale Suche

und Metadaten in einer integrierten Plattform, um Forschern vielseitige Werkzeuge zur Suche nach passendem Probenmaterial zu geben und bei den Biobankern gleichzeitig ein hohes Maß an Datenhoheit zu bewahren.

**Diskussion.** Trotz ihrer Unterschiede schließen sich die vorgestellten Strategien zur Vernetzung von Biomaterialbanken gegenseitig nicht aus. Vielmehr lassen sie sich in gemeinsamen Forschungsinfrastrukturen sinnvoll ergänzen und sie können sogar voneinander profitieren.

### Schlüsselwörter

Verbundforschung · Föderation · Datenintegration · Register · Forschungsinfrastruktur

## Strategies for biobank networks. Classification of different approaches for locating samples and an outlook on the future within the BBMRI-ERIC

### Abstract

**Background.** Medical research projects often require more biological material than can be supplied by a single biobank. For this reason, a multitude of strategies support locating potential research partners with matching material without requiring centralization of sample storage.

**Objectives.** Classification of different strategies for biobank networks, in particular for locating suitable samples. Description of an IT infrastructure combining these strategies.

**Materials and methods.** Existing strategies can be classified according to three criteria: (a) granularity of sample data: coarse bank-

level data (catalogue) vs. fine-granular sample-level data, (b) location of sample data: central (central search service) vs. decentral storage (federated search services), and (c) level of automation: automatic (query-based, federated search service) vs. semi-automatic (inquiry-based, decentral search). All mentioned search services require data integration. Metadata help to overcome semantic heterogeneity.

**Results.** The “Common Service IT” in BBMRI-ERIC (Biobanking and BioMolecular Resources Research Infrastructure) unites a catalogue, the decentral search and metadata in

an integrated platform. As a result, researchers receive versatile tools to search suitable biomaterial, while biobanks retain a high degree of data sovereignty.

**Conclusions.** Despite their differences, the presented strategies for biobank networks do not rule each other out but can complement and even benefit from each other.

### Keywords

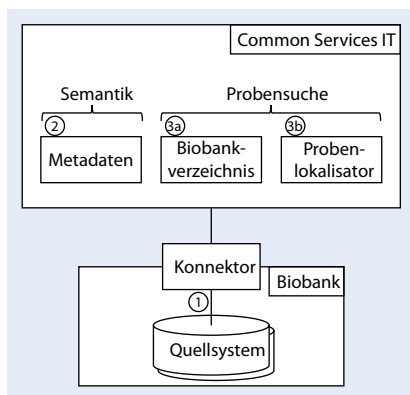
Joint research · Federation · Data integration · Registry · Research infrastructure

scheiden sind: abfragebasierte förderierte Suchdienste (z. B. EHR4CR [17, 18] oder das auf i2b2-basierende SHRINE [19, 20]) antworten dem Forscher automatisch und sofort mit Informationen über verfügbares Biomaterial, z. B. über die Anzahl der in verschiedenen Biobanken jeweils verfügbaren, zu den Suchkriterien passenden Proben.

**Dezentrale Suche.** Die dezentrale Suche basiert hingegen nicht auf einer Abfrage, sondern auf einer Anfrage: Sie holt vor der Rückmeldung eines lokalen Suchergebnisses an die anfragende Such-

komponente die explizite Freigabe durch den jeweiligen Biobankbetreiber ein [21]. Als ergänzende Information für eine solche Freigabeentscheidung enthält jede Suchanfrage in der Regel die Identität des anfragenden Forschers und eine Kurzbeschreibung des geplanten Forschungsprojekts. Technisch setzt dieses Verfahren als lokales Data-Warehouse einen sog. „Brückenkopf“ ein, der Suchantworten erst nach einem interaktiven Dialog mit einem lokalen Verantwortlichen und seiner expliziten Freigabe nach extern übermittelt. Durch die Einbeziehung eines lokalen Biobankverantwortlichen

ergeben sich für die Beantwortung von Suchanfragen zwar deutliche zeitliche Verzögerungen, die sich für einige Anwendungsfälle erschwerend auswirken. Doch wird dadurch die Datenhoheit der Biobank vollständig bewahrt und mit der Vernetzung sind wesentlich niedrigere Hürden im Hinblick auf Datenschutz und Patienteneinwilligungen verbunden, was die Chancen auf eine Teilnahme einer Biobank erhöht [14, 21].



**Abb. 2** ▲ Komponenten des BBMRI-ERIC CS-IT (vereinfacht, eingekreiste Zahlen s. Text)

## Metadaten

Grundlegende Voraussetzung für die im Rahmen von Suchdiensten erforderliche Datenintegration ist ein gemeinsames Verständnis der Biobanken und der Dienste über die semantische Interpretation der probenbeschreibenden Daten. Dieses kann aufgrund unterschiedlicher Dokumentationspraxen der Biobanken nicht vorausgesetzt werden. Die Überwindung dieser semantischen Heterogenität wurde von Riegman et al. [1] als „full-blown annotation problem“ bezeichnet. Es wurde bislang zumeist mittels eines gemeinsam abgesprochenen Datensatzes und einer zugehörigen Vereinheitlichung aller darin enthaltenen Merkmale und deren Ausprägungsmöglichkeiten angegangen. Dieser Datensatz, den alle datenliefernden Biobanken bedienen können müssen, wird also als „kleinster gemeinsamer Nenner“ gebildet und daher auch oft „Minimaldatensatz“ genannt. Um am Suchdienst teilzunehmen, müssen IT-Verantwortliche der Biobanken die jeweils lokal im LIMS vorliegenden Attribute auf die jeweilige Entsprechung in diesem gemeinsamen Datensatz abbilden und beim Export entsprechend umschreiben. Dies ist allerdings nur realisierbar, wenn eine 1:1-Abbildung zwischen Merkmalen sowie auch deren Merkmalsausprägungen möglich ist. Ansätze zum Umgang mit komplexeren Abbildungsszenarien sind z. B. in Mate et al. [22] beschrieben.

Prinzipbedingt problematisch an jedem Minimaldatensatz ist erstens, dass Daten ausschließlich in der konsentierten Datensatzbreite zur Verfügung stehen.

Erfasst also eine Biobank darüber hinausgehende Informationen, stehen diese im Suchdienst nicht zur Verfügung und können somit die Probensuche nicht unterstützen. Der Detailgrad durchsuchbarer Annotationen wird dadurch eingeschränkt. Etwaige Änderungen am Minimaldatensatz sind zwar möglich, aber aufwändig, da nicht nur ein erfüllbarer Konsens unter den beteiligten Biobanken hergestellt werden muss, sondern auch die etablierten Prozesse der Datenintegration entsprechend angepasst werden müssen. Zweitens ist das Format von Minimaldatensätzen nicht standardisiert. Das ist für IT-Systeme ein Problem, da sich die korrekte Interpretation eines Datenelements selten aus seinem Bezeichner allein ableiten lässt. Vielmehr muss jede Annotation einer Bioprobe so umfassend beschrieben sein, dass man damit sowohl den Ursprung als auch die Aussage korrekt herleiten kann. Beispielsweise ist eine prosaische Beschreibung zur Interpretation durch einen menschlichen Leser hilfreich, während Datentypen und Einschränkungen des Wertebereichs IT-Systemen eine Validierung ermöglichen. Probenannotationen sind also wiederum mit Annotationen zu beschreiben. Diese „Annotationen von Annotationen“ werden als Metadaten bezeichnet und können bspw. nach dem ISO-Standard 11179-3:2013 beschrieben werden [23].

Metadaten geben nicht nur dem ETL-Beauftragten eine Interpretationshilfe, sondern auch dem Forscher, der die Daten zur Auswertung erhält. Mit einer maschinenlesbaren Beschreibung und Webschnittstelle kann auch jedes IT-System, das die Daten verarbeitet, Metadaten abrufen und verarbeiten. Zu diesem Zweck erfolgt deren Verwaltung, Pflege und Bereitstellung in einem sog. „Metadata Repository“ (MDR).

Im Sinne einer umfassenden Interoperabilität sollten Metadaten – anders als die Daten selbst – für möglichst viele Systeme direkt abrufbar sein. Das lässt sich einerseits durch eine zentrale Metadatenhaltung erreichen. Andererseits ist auch eine Föderation mehrerer MDR möglich. Einen ersten Schritt hin zu einer solchen MDR-Föderation geht das Samplify MDR des Deutschen Konsortiums für

Translationale Krebsforschung (DKTK, [24]). Das ISO-11179-basierte MDR übernimmt Metadaten aus dem (auf demselben Standard basierenden) caDSR des NCI und reichert sie um für das DKTK wichtige Informationen an (bspw. um eine deutsche Übersetzung [25]). Dabei bleibt der Bezug zu den caDSR-Metadaten erhalten, was einer Fragmentierung entgegenwirkt und die Vergleichbarkeit vereinfacht.

## Biobankvernetzung auf europäischer Ebene

Eine Anwendung der vorgestellten Methoden soll im Rahmen der in BBMRI-ERIC auf europäischer Ebene geplanten Vernetzungsaktivitäten in den kommenden Jahren schrittweise umgesetzt werden. In der BBMRI-Terminologie werden die entsprechenden Softwarelösungen in ihrer Gesamtheit als „Common Service IT“ bezeichnet und im Folgenden auch als CS-IT abgekürzt. Während BBMRI-ERIC und die CS-IT eine Vielzahl von Anwendungsfällen abdecken – bspw. auch die Anforderung und das Tracking von Proben nach erfolgreicher Projektvermittlung – soll hier die Suche nach Proben und passenden Projektpartnern beleuchtet werden.

Die dafür im CS-IT umzusetzende Plattform besteht sowohl aus zentralen Komponenten, die gemeinsam für alle teilnehmenden Biobanken betrieben werden, als auch aus lokalen Komponenten, die jede Biobank unter eigener Hoheit betreibt. Im Einzelnen sind folgende Komponenten vorgesehen (vgl. eingekreiste Zahlen in ■ Abb. 2):

1. Jede Biobank erhält einen sog. *Konnektor*, der im Kontext des oben beschriebenen anfragebasierten, föderierten Suchdienstes einem Brückenkopf entspricht. Er stellt das Bindeglied zwischen den Primärsystemen der Biobank und den zentralen Komponenten des CS-IT dar. Die eigentlichen Primärsysteme (bspw. ein LIMS bzw. Biobank-Management-System) können also unverändert weiterbetrieben werden, müssen aber ihre probenbeschreibenden Daten (bzw. eine Teilmenge davon) auf geeignete



Weise dem Konnektor zur Verfügung stellen. Alternativ könnte zwar die Konnektorschnittstelle auch direkt im Primärsystem implementiert werden. Dies dürfte aber in den wenigsten Fällen realisierbar sein, da dafür meist kommerzielle Softwareanbieter ihre Produkte entsprechend anpassen müssten.

2. *Metadaten* werden zentral durch den CS-IT verwaltet, um die Herstellung semantischer Interoperabilität zwischen den heterogenen Datenbeständen in den Quellsystemen zu unterstützen.
3. Bei der eigentlichen Vernetzung verfolgt der CS-IT zwei Strategien:
  - a. Ein *Biobankverzeichnis* bietet eine Übersicht über alle an den CS-IT angeschlossenen Biobanken. Gemäß der Klassifikation in **Abb. 1** entspricht es einem Katalog, der aggregierte Daten auf Biobankebene enthält, diese anschaulich darstellt und auch einfache Suchabfragen ermöglicht.
  - b. Ein *Probenlokalisator* („sample locator“) implementiert eine dezentrale Suche als Dienst zur Suche nach Daten auf Probenebene. Der Lokalisator ist dabei die Komponente, die dem Forscher die Formulierung von Suchkriterien erlaubt und diese vorhält. Durchgeführt wird die Suche durch den Konnektor in der Biobank unter Einbeziehung eines Biobankverantwortlichen. Eine ausführliche Beschreibung des Prozesses findet sich in [21].

Die Umsetzung und der Betrieb des CS-IT ruht ab Herbst 2015 auf den Schultern von 15 „National Node IT Representatives“ aus den teilnehmenden EU-Staaten.

## Diskussion

Wie die vorgeschlagene Klassifikation zeigt, wird die Probensuche als Teil der Vernetzung von Biomaterialbanken weltweit mit einer Vielzahl von Strategien verfolgt, die sich sowohl in Methodik als auch Zielsetzung wesentlich voneinander unterscheiden. Dennoch schließen sie sich keinesfalls gegenseitig aus,

sondern können vielmehr in geeigneter Kombination zusätzliche positive Synergieeffekte erzeugen. So vereint die dargestellte IT-Plattform von BBMRI-ERIC mit dem Namen „Common Service IT“ – nach Wissen der Autoren erstmals – die Methoden des Katalogs und der dezentralen Suche in einer zusammenhängenden Vernetzungsstrategie. Damit ergeben sich vorteilhafte Synergien gegenüber einer exklusiven Nutzung der einzelnen Methoden. Dies soll an zwei Beispielen illustriert werden:

- Aufgrund der erforderlichen manuellen Dateneingabe sind traditionelle Kataloge möglicherweise nicht immer aktuell. Da im vorliegenden Konzept jedoch zusätzlich eine dezentrale Suche zum Einsatz kommt, ist in teilnehmenden Biobanken ein Brückenkopf vorhanden, der auch eine strukturierte, automatische Meldung von Biobankbeschreibenden und -aggregierten Daten an Kataloge möglich macht. Vorteil ist neben der gesteigerten Aktualität der Daten auch eine geringere Fehleranfälligkeit im Vergleich zur sonst bei Katalogen üblichen manuellen Eingabe im Webbrowser. Insbesondere lassen sich die entsprechend aggregierten Daten auf Basis der implementierten Zugriffsmechanismen im Konnektor in regelmäßigen Abständen direkt ermitteln. Hingegen stellen manuell gemeldete, aggregierte Daten oft nur grobe Schätzungen dar, da sich der Probenumfang einer Biobank täglich ändern kann.
- Die dezentrale Suche kann mithilfe des Katalogs ein höheres Maß an Transparenz herstellen. So kann dem Forscher eine Übersicht darüber gegeben werden, welche Biobanken seine Suchanfrage erhalten werden. Durch die vom Katalog bereitgestellten Daten auf Biobankebene kann diese Übersicht anschaulich und informativ ausfallen; schließlich wurden die Daten im Katalog zu diesem Zweck erhoben.

Über die vorgeschlagene Infrastruktur erhalten Forscher vielseitige Werkzeuge zur Suche nach passendem Biomaterial.

Gleichzeitig wird Biobankern ein hohes Maß an Datenhoheit bewahrt.

## Metadaten und etablierte Standards

Auch Metadaten und (Minimal-)Datensätze etablierter Standards sind – wie die oben vorgestellten Suchmethoden – keine gegensätzlichen Konzepte, sondern können einander ebenfalls ergänzen. So wäre ausgehend von einem leeren Metadata-Repository ein großer Aufwand zu betreiben, um zu dokumentierende Datensätze zu spezifizieren und mit möglichst aussagekräftigen Metadaten zu annotieren. Durch die Aufnahme standardisierter Datensätze, die im besten Fall bereits Bestandteil der ohnehin lokal durchgeführten Dokumentation sind, kann dieser Aufwand reduziert werden. Beispiele solcher standardisierter Merkmale bzw. Datensätze sind a) SPREC („Standard PREanalytical Code“, [26]), ein mehrdimensionaler Code mit Merkmalen, die einen Einfluss auf die Qualität einer Probe und damit auch auf die daraus abgeleiteten Erkenntnisse haben können, und b) der MIABIS-Datensatz („minimum information about biobank data sharing“), der u. a. auch Angaben zur jeweiligen Biobank, zum jeweiligen Patienten sowie zu den pathologischen Befunden umfasst [27].

## Fazit

**Eine Klassifikation von Strategien zur Vernetzung von Biobanken mit Einordnung bestehender Ansätze wurde vorgestellt. Trotz ihrer großen Unterschiede schließen sich ihre Anwendungen gegenseitig nicht grundsätzlich aus. Vielmehr lassen sie sich in Forschungsinfrastrukturen – wie der des europäischen CS-IT aus BBMRI-ERIC – sinnvoll miteinander ergänzen. Als Resultat erhalten Forscher vielseitige Werkzeuge zur Suche nach passendem Biomaterial und Biobanker ein hohes Maß an Datenhoheit.**

## Korrespondenzadresse

### Dr. rer. nat. M. Lablans

Medizinische Informatik in der  
Translationalen Onkologie  
Deutsches Krebsforschungszentrum  
Im Neuenheimer Feld 280, 69120 Heidelberg  
m.lablans@dkfz.de

**Danksagung.** Ein Team, bestehend aus den nationalen IT-Ansprechpartnern der an BBMRI-ERIC teilnehmenden Länder, hat an der Detaillierung und Konzeptionierung dieses Modells mitgewirkt. Besonders hervorzuheben ist die Rolle von Prof. Morris Swertz aus den Niederlanden und Prof. Michael Hummel aus Deutschland.

## Einhaltung ethischer Richtlinien

**Interessenkonflikt.** M. Lablans, D. Kadioglu, S. Mate, I. Leb, H.-U. Prokosch und F. Ückert geben an, dass kein Interessenkonflikt besteht.

Dieser Beitrag enthält keine Studien an Menschen oder Tieren.

**Open Access** Dieser Artikel unterliegt den Bedingungen der Creative Commons Attribution License. Dadurch sind die Nutzung, Verteilung und Reproduktion erlaubt, sofern der/die Originalautor/en und die Quelle angegeben sind.

## Literatur

- Riegman PHJ, Morente MM, Betsou F, Blasio P de, Geary P (2008) Biobanking for better healthcare. *Mol Oncol* 2(3):213–222. doi:10.1016/j.molonc.2008.07.004
- Hummel M, Rufenach C (2015) Biomaterialbanken als Grundlage für die Entwicklung genetisch basierter Präventionskonzepte. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 58(2):127–130. doi:10.1007/s00103-014-2090-5
- Bundesministerium für Bildung und Forschung (2010) Richtlinien zur Förderung einer Nationalen Biomaterialbanken Initiative. <http://www.gesundheitsforschung-bmbf.de/de/2426.php>
- Bundesministerium für Bildung und Forschung (2015) Nationale Biomaterialbanken Initiative. <http://www.gesundheitsforschung-bmbf.de/de/2638.php>
- Revermann C, Sauter A (2007) Biobanken als Ressource der Humanmedizin. Bedeutung, Nutzen, Rahmenbedingungen. *Studien des Büros für Technikfolgen-Abschätzung beim Deutschen Bundestag*, Bd 23. Edition Sigma, Berlin. ISBN 9783836081238
- Webseite des German Biobank Node (GBN). <http://www.bbMRI.de/>
- Webseite von BBMRI-ERIC. <http://www.bbMRI-eric.eu/>
- Demchok J, Taube S, Fombonne B, Lubensky I (2013) The National Cancer Institute (NCI) specimen resource locator. *Eur J Cancer* 49:S37–S38. doi:10.1016/S0959-8049(13)70196-5
- TMF e. V. Webseite des Deutschen Biobanken-Registers. <http://www.biobanken.de/>
- Wichmann H, Kuhn KA, Waldenberger M et al (2011) Comprehensive catalog of European biobanks. *Nat Biotechnol* 29(9):795–797. doi:10.1038/nbt.1958
- TMF e. V (2012) Ein Schaufenster für medizinische Biobanken. <http://www.biobanken.de/News/tabid/83/articleType/ArticleView/articleId/22/language/de-DE/Ein-Schaufenster-fur-medizinische-Biobanken.aspx>
- TMF e. V (2015) Biobank registries in Europe struggle with similar challenges. <http://www.tmf-ev.de/News/articleType/ArticleView/articleId/1698.aspx>
- Schmelcher DK (2013) Integrationskonzepte und -lösungen zur Etablierung einer Forschungsinfrastruktur für Biobanken. Dissertation. Technische Universität, München. <https://mediatum.ub.tum.de/doc/1129699/1129699.pdf>
- Lablans M (2015) Die dezentrale Suche in der medizinischen Verbundforschung. Unv. Dissertation, Westfälische Wilhelms-Universität, Münster
- Schröder C, Heidtke KR, Zacherl N, Zatloukal K, Taupitz J (2011) Safeguarding donors' personal rights and biobank autonomy in biobank networks: the CRIP privacy regime. *Cell Tissue Bank* 12(3):233–240. doi:10.1007/s10561-010-9190-8
- Leser U, Naumann F (2007) Informationsintegration. Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen. dpunkt-Verlag, Heidelberg. ISBN 9783898644006
- Doods J, Bache R, McGilchrist M, Daniel C, Dugas M, Fritz F (2014) Piloting the EHR4CR feasibility platform across Europe. *Methods Inf Med* 53(4):264–268. doi:10.3414/ME13-01-0134
- Ouagne D, Hussain S, Sadou E, Jaulent M, Daniel C (2012) The Electronic Healthcare Record for Clinical Research (EHR4CR) information model and terminology. *Stud Health Technol Inform* 180:534–538. doi:10.3233/978-1-61499-101-4-534
- McMurry AJ, Murphy SN, MacFadden D et al (2013) SHRINE: enabling nationally scalable multi-site disease studies. *PLoS One* 8(3):e55811. doi:10.1371/journal.pone.0055811
- Weber GM, Murphy SN, McMurry AJ et al (2009) The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc* 16(5):624–630. doi:10.1197/jamia.M3191
- Lablans M, Kadioglu D, Muscholl M, Ückert F (2015) Exploiting distributed, heterogeneous and sensitive data stocks while maintaining the owner's data sovereignty. *Methods Inf Med* 54(4):346–352. doi:10.3414/ME14-01-0137
- Mate S, Köpcke F, Toddenroth D et al (2015) Ontology-based data integration between clinical and research systems. *PLoS One* 10(1):e0116656. doi:10.1371/journal.pone.0116656
- ISO/IEC JTC1 SC32 WG2 (2014) ISO/IEC 11179 Information Technology – Metadata registries. <http://metadata-standards.org/11179/>
- Lablans M (2012) Konzept der CCP-IT des DKTK. <http://ccp-it.dkkt.dkfz.de>
- Warzel DB, Andonaydis C, McCurry B, Chilukuri R, Ishmukhamedov S, Covitz P (2003) Common data element (CDE) management and deployment in clinical trials. *AMIA Annu Symp Proc* 2003:1048
- Lehmann S, Guadagni F, Moore H et al (2012) Standard preanalytical coding for biospecimens: review and implementation of the Sample PREanalytical Code (SPREC). *Biopreserv Biobank* 10(4):366–374. doi:10.1089/bio.2012.0012
- Norlin L, Fransson MN, Eriksson M et al (2012) A minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreserv Biobank* 10(4):343–348. doi:10.1089/bio.2012.0003